DOCUMENT RESUME

ED 482 078 IR 058 764

AUTHOR	Perkins, John
TITLE	A New Way of Making Cultural Information Resources Visible on the Web: Museums and the Open Archive Initiative.
PUB DATE	2001-00-00
NOTE	8p.; In: Museums and the Web 2001: Selected Papers from an International Conference (5th, Seattle, Washington, March 15-17, 2001); see IR 058 756.
AVAILABLE FROM	Archives & Museum Informatics, 2008 Murray Ave., Suite D, Pittsburgh, PA 15217; e-mail: info@archimuse.com; Web site: http://www.archimuse.com/. For full text: http://www.archimuse.com/mw2001/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Access to Information; *Archives; Information Dissemination; *Information Management; Information Sources; Information Storage; Metadata; *Museums; Online Systems; World Wide Web
IDENTIFIERS	Dublin Core; *Museum Collections

ABSTRACT

Museums hold enormous amounts of information in collections management systems and publish academic and scholarly research in print journals, exhibition catalogs, virtual museum presentations, and community publications. Much of this rich content is unavailable to web search engines or otherwise gets lost in the vastness of the World Wide Web. The Open Archives Initiative (OAI) has developed an easily implemented protocol to enable data providers to expose their information and service providers to access and use it. The CIMI Consortium is working with the OAI to make it possible for museums to enhance the availability of their research resources, allowing them to be discovered in Web-space by the specialist audiences for which they are intended or by service providers who collect, distribute or in other ways provide access. By building on the OAI protocol, Dublin Core, and museum community XML developments, significant advancements can be made in exposing museum information resources. This paper introduces the OAI and its protocol, explores its potential relevance to museums, presents CIMI's work as an alpha tester of OAI, and looks ahead to future developments. (Contains 11 references.) (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Bearman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 482 078

Register
Workshops
Sessions
Speakers
Interactions
Demonstrations
Exhibits
Events
Best of the Web
Key Dates
Seattle
Sponsors

A&MI
Archives & Museum Informatics
2008 Murray Ave
Suite D
Pittsburgh, PA
15217 USA
info@archimuse.com
www.archimuse.com

Search A&MI

Join our Mailing List

Privacy

Updated: 02/27/2001 11:24:44

PAPERS
Museums and the Web 2001

A New Way Of Making Cultural Information Resources Visible On The Web: Museums And The Open Archives Initiative

John Perkins, CIMI Consortium, Canada

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

Museums hold enormous amounts of information in collections management systems and publish academic and scholarly research in print journals, exhibition catalogues, virtual museum presentations, and community publications. Much of this rich content is unavailable to web search engines or otherwise gets lost in the vastness of the World Wide Web. The Open Archives Initiative (OAI) has developed an easily implemented protocol to enable data providers to expose their information and service providers to access and use it. The CIMI Consortium is working with the OAI to make it possible for museums to enhance the availability of their research resources, allowing them to be discovered in Web-space by the specialist audiences for which they are intended or by service providers who collect, distribute or in other ways provide access. By building on the OAI protocol, Dublin Core, and museum community XML developments, significant advancements can be made in exposing museum information resources. This paper introduces the OAI and its protocol, explores its potential relevance to museums, presents CIMI's work as an alpha tester of OAI, and looks ahead to future developments.

Keywords: Open Archives Initiative, OAI, CIMI, metadata, metadata harvesting, Dublin Core, MXL

Introduction

The ubiquity of the Web and success of popular search engines have fueled an expectation for quick, easy, and successful results in the quest for information and knowledge. Increasingly, scholars, students and other explorers are turning to the Web for their research needs and relying less often on traditional research sources. Museums have immensely rich information resources in publications, research papers, exhibition catalogues, virtual museums, databases, and intranets, but access to much information of value about the kinds of materials museums hold is rarely available through web search engines. Internet search engines only reach static HTML web pages, but much of what museums have is opaque to the indexers because it is in databases, dynamically generated, or in some other non-HTML form. These resources constitute what is becoming known as the hidden Web, estimated to contain 400-550 times more content than the commonly defined Web. (BrightPlanet 2001)

IR058764

BEST COPY AVAILABLE

If this problem alone were solved and all the hidden web resources were suddenly available for indexing, the difficulty of finding reliable, useful, precise information would be seriously compounded, not alleviated. One way to address this is through collecting and indexing metadata records, rather than indexing the entire contents of HTML pages, thereby providing greater possibilities for precision. This is essentially the traditional library approach of creating descriptive metadata and building union catalogues. However, library catalogues are expensive to maintain and in the Web world, both difficult to find and hard to search across.

As separate approaches, it seems neither the old library methods nor the new Internet approach is serving researchers and scholars particularly well. (CLIR 2001)

A particularly promising solution is to explore the utility of combining the best of traditional library and museum techniques, such as creating descriptive metadata records in catalogues, with the best of new Internet techniques like large scale, machine harvesting of information. It is possible to consider this because of new developments in Web workable technical protocols, the uptake of XML as a way to package and transfer information, and the development of international standards for describing museum metadata content.

The Open Archives Initiative

The Open Archives Initiative, OAI, (http://www.openarchives.org) develops and promotes technical protocols and standards, collectively called the OAI technical framework, to facilitate access to scholarly research information on the Web. It is based on the premise that a simple, easily implemented technical framework can allow holders of information to create repositories of metadata describing their resources that in turn can be harvested and made available for further processing or use. (OAI Protocol 2001)

The OAI technical framework describes how *repositories* of metadata about information resources are constructed. Repositories are essentially network accessible servers offered by data providers. A repository makes available via a simple protocol *records* that contain metadata about its items (content). A repository may, optionally, organize its items into sets corresponding to its collections or other groups, thus allowing clients to harvest metadata records selectively.

A *record* is an XML encoded byte-stream that serves as a packaging mechanism for harvested metadata. The OAI protocol mandates the use of unqualified Dublin Core as the common record for discovery. (Dublin Core 2001) It also allows community-specific metadata sets described by XML SCHEMAS for more detailed description based on the assertion that both simple metadata for interoperability and cross-domain discovery as well as a method for conveying richer community-specific descriptions are needed.

All OAI repositories must recognize a set of requests or *verbs* carried in http POST or Get methods that allow access to the metadata records. It is through these commands that metadata is harvested and transferred.

One design criteria of the OAI technical framework of particular relevance to individual communities such as the museum community is

the notion of *extension packages*. Not only does the protocol allow a community to expose its own metadata schema, but it also allows other extensions such as unique collection level metadata or, if deemed necessary, rights metadata. The OAI protocol doesn't place limits on the number of allowable metadata sets, but does specify that their data formats be describable by an XML Schema.

In order to federate distributed repositories, the OAI has established a registry service available through the OAI web page to provide a list of publicly available repositories and to provide a mechanism for conformance testing. (OAI Registry 2001)

The potential of OAI technical framework is in providing the enabling technology for the federating of distributed information resources and their discovery and use. The power of the OAI technical framework is in its simplicity and ease of implementation.

Describing Information Resources for Discovery: Dublin Core and XML

While the OAI protocol defines new technical standards for repositories and the machine-to-machine dialogue between data providers and harvesters, it draws on the established international standard Dublin Core for the mandatory metadata record format. (Dublin Core 2001) The Dublin Core metadata set was developed specifically to allow a simple and easy-to-use description of information resources for their discovery. The utility of Dublin Core was corroborated by CIMI in its Dublin Core Metadata Testbed that explored the use of unqualified Dublin Core for discovery of museum resources, both at a coarse grain level and at a more detailed, complex level. At the higher, coarse grain level, the Dublin Core is effective both for discovery of resources and as a means for museums to interoperate with other communities in a networked environment such as the World Wide Web. (CIMI 1999a)

To go beyond simple discovery and interoperation, the OAI anticipated, through inclusion of the *extension packages* concept, that in addition to a core metadata format, individual communities of implementers would require additional descriptive formats. Again, this need was borne out in the CIMI Dublin Core testbed findings where it was concluded that extending the Dublin Core to handle community-specific needs was problematic. (CIMI 1999a)

Alternatives need to be found to extending or qualifying Dublin Core to facilitate the more complete descriptions needed by the museum community. The OAI addresses this by allowing support for parallel metadata sets. For museums, this could conceivably include record structures such as SPECTRUM (rich museum object information), CIMI (public access), AMICO (art museum images), MIDIS (monuments and built environment), OBJECT ID (loss and theft), and RLG Inc.'s CMI (Cultural materials).

The challenge is that each community of OAI implementers must agree on what metadata formats are needed beyond the core, and must provide XML SCHEMAS for each of them. Once this is accomplished, the metadata foundations will be in place for use of the OAI protocol.

Early in the development of the OAI, CIMI recognized it had a number of

features that could help significantly advance access to museum information. First and perhaps most importantly, the OAI protocol was simple and appeared to be easy to implement using tools and skills (Webservers, http, JAVA, PERL, CGI etc.) within the easy reach of museums. Secondly, it relied on the Dublin Core as a metadata format for the simple discovery of information resources within and between communities. This format was proven workable for museums, and there exists a guide to best practice for its use. (CIMI 1999b) Finally, the OAI mandated XML for packaging richer metadata sets and transferring records. XML is a standard that is gaining wide acceptance in museums, and XML SCHEMAS exist or are in the process of being created for many of the community standards mentioned above.

CIMI's test of OAI V.1.0

Because of the perceived potential of OAI for museums, CIMI participated as a pre-release tester of the OAI protocol. (OAI Alpha Test 2001) As part of the test, we built a generic OAI-compliant repository. (CIMI OAI Repository 2001). The repository architecture shown in Figure 1 uses a layered approach, standardized APIs, a generic http interface, and interchangeable components. This allows implementers the use of different back-end databases, webservers, or XML generators and minimizes hard-wired coding.

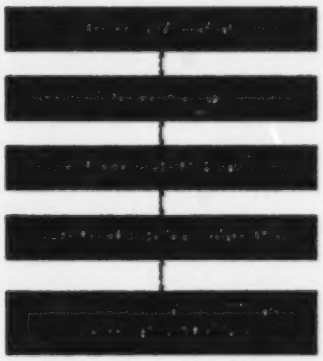

Figure 1: CIMI OAI Repository Layered Architecture

The repository took a skilled JAVA programmer two weeks elapsed time to build. This period included both an orientation to CIMI and the OAI as well as reading and understanding the protocol, and then building the application. The development process started with designing a JAVA API for the repository and a JAVA servlet to interface between http/OAI protocol layers and the repository. The reference repository was written using MySQL and JDBC.

The CIMI reference application serves Dublin Core records from an

BEST COPY AVAILABLE

Apache Webserver generated by the earlier CIMI testbed from the MySQL database. Because of the modularity inherent in the architecture, the Repository could be layered on top of any ODBC-compliant database, be served from other servers, and make use of different XML generators.

Looking Ahead

The initial evaluation demonstrated that the OAI protocol is indeed simple to build. CIMI has limited technical resources and skills but was nonetheless able to successfully build an OAI repository that appears to be useful. Based on the positive experience as an alpha implementer, CIMI plans to continue explorations of the OAI protocol and research its use by museums.

One way is by making the code for the CIMI repository and its associated explanatory materials available for downloading from the CIMI Website. (CIMI Publications 2001) We hope museums will take advantage of its availability to install, experiment with and use the protocol. We hope to compile and report the experiences of these *ad hoc* tests.

CIMI is also interested in conducting a more formal, large-scale test of the OAI for museums as a CIMI testbed. As part of this work, we propose using OAI V.1.x in combination with scoped extensions and other applications necessary for aggregation processes (e.g. editorial control, content management and enhancement, registry) to harvest and collect museum metadata from cultural memory organizations. It will focus on materials that document culture and civilizations, including museum objects, art, images, and related materials. We will structure this as a CIMI testbed, inviting participation from a group of interested members. We expect respondents to include national museum organizations, individual museums, commercial enterprises, and museum system vendors. Once underway, the project will run 12-18 months in concert with projects in other communities and the OAI test period.

The purpose of the research is to explore how a specific community of users can use the OAI protocol. Part of this is to investigate what agreements users need to make within the protocol framework itself (e.g. additional metadata sets), and part is to identify any extensions or modification required to make the framework additionally useful. Our testbed will give museums a place to expose their metadata and promote their institutions, test the OAI protocol for utility in describing non-bibliographic resources, and could provide a rich resource of cultural metadata leading eventually to the materials themselves and the institutions offering them.

It is one thing to test the technical viability of the OAI protocol by implementing the protocol at a technical level, but another to imagine and determine useful services that might be built on it. We have imagined a number of scenarios that could be tested.

We imagine, for example, that services like AMOL (Australian Museums Online), AMICO, the Canadian Digital Museum, or RLG Inc.'s Cultural Materials Initiative might want to add a feature to "search for more like this" in collections or repositories not under their direct control. We imagine that individual museums or groups of museums all using the

same collections management system might make use of the repository for internal operational needs, for scholarly access, as well as for supplementing information services they provide publicly. We imagine that commercial services such as AskArt (http://www.askart.com) - a directory of American Artists - or Virtualology (http://www.virtualology.com) - a virtual education project - would find the resource attractive and useful. We imagine that an easy-to-use protocol might be attractive to sales and auction houses, encouraging them to make useful research information resources available (such as those now manually compiled). We expect national service providers like the UK JISC higher education information services to have an interest in using museum repositories. We know that the operators of the new Internet top-level-domain for museums (MusDoma) are extremely interested in providing directory-like services that would include search access to our harvested cultural materials metadata. We also imagine that harvesting exhibition catalogues and museum publications from library catalogues, artist biographies, museological literature from A&I services, and sales records from auction houses is of interest to museum researchers. These all are the kinds of services that might emerge once the OAI is widely deployed in the museum community.

Regardless of the services developed, there will be a number of issues relating to widespread adoption of the OAI protocol in the museum community. We foresee a need for our community to test hypotheses, assertions, and issues such as:

- the utility of the Dublin Core for meeting information requirements of service providers and consumers;
- the functionality of the OAI protocol as a basis for a harvesting service, including issues of hierarchical descriptions, scalability, required extensions, presentation and partitioning;
- community extensions required for the OAI and DC in order to provide useful metadata within and between communities;
- requirements and practices for content management, metadata enhancement, and editorial control
- aggregation, integration, access and presentation of bibliographic, textual, multiple media, and object metadata
- the need, scope and services of a registry
- access control and rights
- mechanisms, processes for paths to underlying content
- business model for sustainability

Conclusion

Both CIMI and many of our members have significant experience in the metadata harvesting business. It is this experience that motivates us to explore the OAI protocol as an enabling technology to facilitate access to resources by making it easier for museums to expose and collect metadata. The OAI protocol in concert with a museum testbed seems a logical and sensible research initiative that will bring us closer to making the rich information resources museums hold more widely available to researchers and other users.

References

BrightPlanet 2001: A description of the Deep Web is at http://www.completeplanet.com/Tutorials/DeepWeb/summary03.asp

CIMI 1999a: CIMI's report of the DC testbed
http://www.cimi.org/publications.html#dc_2

CIMI 1999b: CIMI's Guide to Best Practice for the use of Dublin Core in museums is available from
http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf

CIMI OAI Repository 2001: OAI compliant harvesters can access the CIMI Repository at: http://www.cimi.org/oai/OAI_test.html. There is a human viewable interface page at the same location.

CIMI OAI 2001: General information about CIMI's work on the OAI can be found at: http://www.cimi.org/oai/index.html

CIMI Publications 2001: The CIMI OAI Repository code and associated explanatory documents can be found in the Metadata Harvesting section of the Publications area on the CIMI Website at:
http://www.cimi.org/publications.html

CLIR 2001: Vision paper for new Web searching methods at
http://www.clir.org/diglib/architectures/vision.htm

Dublin Core 2001: Detailed information about the Dublin Core, its activities and metadata sets can be found at: http://www.dublincore.org

OAI Registry 2001: At the time of writing the OAI registry was still being constructed. By the time of publication access to the online registration should be at http://www.openarchives.org

OAI Alpha Test 2001: A listing of participants and their experiences is found at: http://www.openarchives.org/OAISC/alpha-testing-press-release.htm

OAI Protocol 2001: The current version of the OAI protocol document is available at http://www.openarchives.org/OAI/openarchivesprotocol.htm

Acknowledgements

The author would like to thank Carl Lagoze for his contributions to the OAI sections of this paper and Henry Stern for developing the CIMI Repository. Thanks are also due to the Open Archives Initiative, the Digital Library Federation, the Mellon Foundation, and the CIMI members for enabling the author to participate in the development of CIMI's thinking on the use of OAI in museums.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").